УДК 534.2: 534.3: 534.4

М.А.СУМБАТЯН, С. Е.ШЕВЦОВ

АЛГОРИТМ ЦИФРОВОЙ ОБРАБОТКИ АКУСТИЧЕСКИХ СИГНАЛОВ АУДИОФАЙЛОВ И ИХ РАСПОЗНАВАНИЕ НА ОСНОВЕ ОБЪЕКТИВНЫХ КРИТЕРИЕВ

Разрабатываются алгоритмы цифровой обработки акустических сигналов, характерных для аудиофрагментов, записанных в виде wav-файлов. Вводится понятие автокорреляционной функции для таких сигналов, а также указываются некоторые объективные параметры, характеризующие сигналы такого типа. Исследуются некоторые практические аспекты в приложении к возможной классификации различных аудиофрагментов.

Ключевые слова: *Цифровая обработка, акустические сигналы, автокорреляционная функция.*

Введение. Стандартные файлы, содержащие аудиофрагменты, обладают рядом специфических особенностей. С одной стороны, они содержат информацию об акустических сигналах, записанных через микрофон, и, следовательно, задача обработки таких сигналов может быть решена классическими методами цифровой обработки [1]. С другой стороны, специфика данных акустических сигналов, связанных со звучанием фрагментов, рассчитанных на восприятие слушателем, отличает их, например, от сигналов акустической эмиссии или каких-либо подобных акустических сигналов, имеющих совершенно иную физическую природу [2].

В связи с этим обработка исследуемых аудиосигналов должна основываться на особых принципах, пригодных для сигналов, предназначенных для прослушивания. В данной работе предлагается использовать оценку поведения автокорреляционной функции (АКФ). Оказывается, что некоторые стохастические параметры поведения этой функции напрямую связаны с типом аудиофрагментов, записанных в виде акустических сигналов в рассматриваемых wav-файлах.

Фильтрация по октавным полосам. Пусть в стандартном wav-файле записан некий аудиофрагмент. Информация о нем содержится в файле в виде некоторого акустического сигнала $p(t)$, характеризующего амплитудно-временную зависимость акустического давления от времени. Современный PCM стандарт подразумевает в общем случае 16-битовую стереозапись с частотой дискретизации 44100 Гц. Это означает, что на самом деле записанный сигнал имеет два канала, в каждом из них – с записью в дискретном виде:

$$p_j = p(t_j),\ t_j = jh_t,\ h_t = 1/44100\ сек.,\ j = 0,...,N-1,\ N h_t = T. \quad (1)$$

В случае стереосигнала функция вида (1) должна рассматриваться для левого и правого акустического каналов отдельно. PCM стандарт допускает запись также и в монорежиме. В этом случае имеем одну функцию вида (1). 16-битовая запись предполагает, что один бит для значений функции p_j отводится под знак, а оставшиеся 15 бит выделяются под значения модуля самой величины в дискретизации данных от 1 до $2^{15} = 32768$.

Таким образом, относительная погрешность в дискретном задании функции равна $1/32768 = 0.003\%$, что следует признать довольно точным приближением для непрерывной функции даже со значительными градиентами.

В классической акустике [3] общепринятым является обработка сигналов в каждой из шести основных октавных полос, в которых частоты колебания выражены в герцах:

$$f^{(1)} = 125,\ f^{(2)} = 250,\ f^{(3)} = 500,\ f^{(4)} = 1000,\ f^{(5)} = 2000,\ f^{(6)} = 4000\,. \qquad (2)$$

При этом нижняя и верхняя границы каждой октавной полосы (соответственно частоты $f_1^{(n)}, f_2^{(n)}$) выражены через несущую частоту соответствующей октавы

$$f_1^{(n)} = f^{(n)}/\sqrt{2}, \qquad f_2^{(n)} = f^{(n)}\sqrt{2}, \qquad (n = 1,\ldots,6), \qquad (3)$$

покрывая непрерывным образом частотную область от $125/\sqrt{2} = 88\,Гц$ до $4000\sqrt{2} = 5657\,Гц$.

Первая часть вычислений, применяемая к записанному сигналу для оценки его октавных компонент на частотных интервалах $(f_1^{(n)}, f_2^{(n)}),\ (n = 1,\ldots,6)$, является достаточно стандартной и основывается на теории цифровых фильтров [4]. Данный метод в применении к произвольному сигналу $p(t)$ при переходе в частотную область состоит в отсечении вкладов от всех частот вне интервала (f_1, f_2), на котором происходит фильтрация. Введем пробную функцию $g(t)$, преобразование Фурье которой равно нулю: $G(f) = 0$ для всех частот вне интервала (f_1, f_2) и равно единице: $G(f) = 1$ для $f \in (f_1, f_2)$. Поскольку каждая частотная компонента f дает вклад в осцилляции на рассматриваемой частоте как со значением f, так и $-f$, то $(\omega = 2\pi f)$

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(f)e^{-i\omega t}d\omega = \frac{1}{2\pi}\int_{f_1<|f|<f_2} G(f)e^{-i\omega t}d\omega =$$
$$= \frac{1}{2\pi}\int_{f_1<|f|<f_2} e^{-i\omega t}d\omega = \frac{1}{\pi}\frac{\sin(\omega_2 t) - \sin(\omega_1 t)}{t} = \frac{\sin(2\pi f_2 t) - \sin(2\pi f_1 t)}{\pi t}\,. \qquad (4)$$

Тогда, согласно теореме о свертке, сигнал, отфильтрованный на частотном отрезке (f_1, f_2), может быть представлен в континуальном и дискретном виде:

$$p^*(t) = \int_{-\infty}^{\infty} p(\tau)g(t-\tau)\,d\tau\,, \qquad p^*(t) = h_t \sum_{j=0}^{N-1} p(t_j)g(t-t_j). \qquad (5)$$

Такой прямолинейный подход обладает одним дефектом, известным в литературе как «эффект Гиббса» [5]. Он обычно наблюдается при применении дискретного аналога преобразования Фурье (ДПФ) вместо его исходной континуальной формы и состоит в том, что для непериодических сигналов ДПФ, как правило, генерирует небольшой сильно осциллирующий «хвост» на концах рассматриваемого временного интервала. Для устранения этих фиктивных осцилляций можно применять те или иные «временные окна», или взвешивание. В нашем алгоритме мы используем окна Хэмминга [5].

Заметим, что такой расчет, осуществленный прямо во временной области, требует квадратичного числа арифметических операций по числу выбранных временных узлов, поскольку в формуле (5) для каждого момента $t = t_m$ необходимо осуществить суммирование по j. Для существенного сокращения времени вычислений необходимо перейти в спектральную область, воспользовавшись теоремой свертки:

$$\tilde{p}^*(\omega) = \tilde{p}(\omega) G(\omega). \quad (6)$$

Такой подход, с использованием быстрого преобразования Фурье (БПФ) в дискретном виде является линейно-логарифмическим, т.е. линейным по числу узлов. Далее, Фурье-обращение формулы (6) – опять-таки линейно-логарифмическим. Однако на пути применения БПФ существует один существенный момент, преодоление которого требует определенного навыка. Дело в том, что по своей сути выражение (5) является вещественным, однако, переход в частотную область (6), без предварительной обработки, приводит к оперированию с комплекснозначными выражениями. Тогда и полученный результат будет комплекснозначным. Для преодоления указанной трудности запишем обращение Фурье выражения (6):

$$p^*(t) = \frac{1}{2\pi}\int_0^{\infty}\left[\tilde{p}_{re}(\omega) + i\tilde{p}_{im}(\omega)\right] G(\omega) e^{-i\omega t} d\omega +$$
$$+\frac{1}{2\pi}\int_0^{\infty}\left[\tilde{p}_{re}(\omega) - i\tilde{p}_{im}(\omega)\right] G(\omega) e^{i\omega t} d\omega = \frac{1}{\pi}\mathrm{Re}\int_0^{\infty}\tilde{p}(\omega) G(\omega) e^{-i\omega t} d\omega, \quad (7)$$

где мы учли, что функция $G(\omega)$ по своему определению вещественна; для вещественной функции $p(t)$ реальная часть ее Фурье образа – функция $\tilde{p}_{re}(\omega)$ четная по ω, а мнимая часть $\tilde{p}_{im}(\omega)$ нечетная по ω. Кроме того, функция $G(\omega)$ четная. Таким образом, искомый результат получается взятием вещественной части от обратного преобразования Фурье в положительной частотной области.

Переход в частотную область часто также сопровождается учетом реальной чувствительности слуховых каналов, которая является не постоянной по частоте. Максимум чувствительности находится примерно в области 2000 - 2500 Гц и по традиции описывается А-взвешенной частотной зависимостью [3]. Проведенные авторами исследования показали, что с достаточно высокой точностью данная функция может быть приближена выражением для энергетической характеристики, связанной с квадратом акустического давления:

$$A^2(\omega) = 10^{y/10}, \quad y = 2 - 50\left(\frac{\lg(f/2200)}{\lg(20/2200)}\right)^2. \quad (8)$$

Характерные точки, через которые проходит график данной функции: $A(20) = -48dB, A(200) = -11dB, A(2200) = 2dB, A(10000) = -3dB.$

Таким образом, вместо функции фильтрации $G(\omega)$ в частотной области на практике следует использовать произведение функций $G(\omega)$ и $A(\omega)$.

Автокорреляционная функция сигнала. После того как исходный аудиосигнал разложен на 6 основных октавных полос, а также учтена чув-

ствительность слуховых каналов (8), в каждой из этих полос происходит независимая цифровая обработка. Исследования Андо [6], выполненные им для всего сигнала в целом, без разложения на частотные интервалы, показали, что вид аудиофрагмента напрямую связан с АКФ записанного сигнала $p(t)$. Данная функция была предложена для описания поведения стационарных процессов [1] и имеет вид

$$\varphi_p(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T p(t)p(t+\tau)\,dt\,. \qquad (9)$$

Оказывается, что для эргодических стационарных процессов АКФ убывает с ростом τ : $\lim_{\tau\to\infty} \varphi_p(\tau) = 0$. Андо [6] показал, что характерная скорость спадания модуля этой функции с ростом аргумента τ однозначно характеризует тип записанного аудиофрагмента. В частности, решающим является значение основного параметра τ_e, определяющего время спадания функции $\varphi_p(\tau)$ на 10 дБ. Исследования авторов показали, что эту идею Андо можно усилить, перенеся вычисление параметра τ_e на различные частотные области, т.е. применяя в (9) вместо функции $p(t)$ функцию $p^*(t)$. Прикладные аспекты будут описаны в следующем пункте. Здесь же опишем алгоритм эффективного вычисления АКФ. Как и в проведении фильтрации по формуле (5), нахождение АКФ-функции, согласно ее определению (9) требует квадратичного числа арифметических операций. Для возможности использования БПФ следует воспользоваться теоремой Винера-Хинчина [1], выражающей АКФ в виде обратного преобразования Фурье от спектральной плотности мощности $\Phi(\omega)$:

$$\varphi_p(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(\omega) e^{-i\omega\tau}\,d\omega\,, \quad \Phi(\omega) = \lim_{T\to\infty} \frac{1}{T} \left| \int_0^T p(t) e^{i\omega t}\,dt \right|^2. \qquad (10)$$

Ясно, что обе интегральные операции в (10) в их дискретной реализации допускают применение БПФ, т.е. такой подход требует линейно-логарифмического числа арифметических операций.

Дискретная реализация описанного алгоритма основана на интерполяционной формуле для интеграла Фурье в (10):

$$\left. \int_0^T p(t) e^{i\omega t}\,dt \right|_{\omega=\omega_k} \approx h_t \sum_{j=0}^{N-1} p_j e^{i\omega_k t_j} = h_t \sum_{j=0}^{N-1} p_j e^{2\pi i k j h_t h_f} = h_t \sum_{j=0}^{N-1} p_j e^{2\pi i k j h_f T/N} =$$
$$= h_t \sum_{j=0}^{N-1} p_j e^{2\pi i k j/N},\ h_f = 1/T = 1/(N h_t),\ \omega_k = 2\pi k h_f,\ \Omega = \omega_{\max} = N h_f. \qquad (11)$$

В такой интерпретации соотношения (10) допускают прямое применение БПФ. При этом сама АКФ определяется на основе обратного ДПФ:

$$\varphi_p(t_j) = \frac{1}{\pi} \mathrm{Re} \int_0^{\Omega} \Phi(\omega) e^{-i\omega t}\,d\omega \approx$$
$$\approx \frac{2\pi h_f}{\pi} \mathrm{Re} \sum_{k=0}^{N-1} \Phi(\omega_k) e^{-i\omega_k t_j} = 2h_f\, \mathrm{Re} \sum_{k=0}^{N-1} \Phi(\omega_k) e^{-2\pi i k j/N}\,, \qquad (12)$$

поскольку $\Phi(\omega)$ - четная вещественная функция. При таком подходе выражение (12) также допускает прямое применение БПФ.

Роль параметра τ_e в ряду объективных акустических параметров. Как было сказано выше, параметр τ_e характеризует эффективный срок нормализованной АКФ, определяемый по промежутку времени, на котором АКФ спадает на 10 дБ. Значение этого параметра для оценки типа аудиофрагмента и качества звучания исследовалось на сигналах записанных фрагментов четырнадцати произведений, исполненных разными составами. Первоначально для определения оптимального значения τ_e в трёх частотных полосах были исследованы диапазоны инструментов, относящихся к трём основным регистрам в музыке и звукотехнике. Так, у контрабаса предел в верхнем диапазоне - это «ля» первой октавы (440 Гц). Однако необходимо было определить его наиболее часто используемый диапазон игры. Для этого был разобран ряд партитур из списка хрестоматийных произведений. Выявлено, что частота раздела между нижней и средней полосой находится в пределах 260-280 Гц, а между средней и верхней – 1200-1300 Гц. Поэтому для дальнейшего исследования авторам пришлось отказаться от двух общепринятых подходов. С одной стороны, Андо [6] оперировал этим параметром для всего сигнала в целом, без разделения на различные частотные диапазоны. С другой стороны, традиционное разделение на октавные полосы, в силу сказанного выше, также неприемлемо для наших экспериментов. В связи с этим для оценки параметра авторами предлагается разделение на три частотные полосы τ_e: низко-, средне-, и высокочастотную, с границами разделения 261 и 1244Гц. В нашем алгоритме с предварительным использованием соответствующих фильтров была проведена работа по нахождению параметра τ_e как в трёх частотных полосах (с соответствующим присвоением символов – hi, mid, low), так и без фильтрации, используя сэмплированные фрагменты. В таблице представлены данные минимальных значений параметра, поскольку именно они представляют интерес, обусловленный большей вероятностью возникновения эффекта – «эхо» в концертном зале в этом диапазоне значений. Выбирались характерные фрагменты музыкальных сигналов (смена темпа, характера, фактуры, вступление других инструментов).

Сигналы музыкальных произведений и их протяжённость в секундах	Исходный импульс τ_e, мс	Высокие частоты τ_e, мс	Средние частоты τ_e, мс	Низкие частоты τ_e, мс
1	2	3	4	5
1. 99	40/41	20/20	54/61	109/113
2. 100	18/27	12/18	28/36	72/54
3. 30	30/29	28/21	51/52	91/93
4. 66	20/32	13/13	43/46	92/85
5. 51	33/40	20/20	40/50	226/123

Окончание табл.

1	2	3	4	5
6. 29	40/40	32/32	40/40	111/97
7. 29	22/30	15/18	34/34	61/68
8. 50	35/32	60/63	140/132	220/215
9. 63	34/25	13/13	34/30	117/118
10. 44	167/139	46/151	167/157	153/123
11. 30	46/46	11/11	91/55	147/183
12. 29	51/51	24/24	51/86	173/194
13. 45	19/19	9/12	37/38	45/136
14. 28	18/18	9/9	34/34	85/85

Примечание: 2,3,4,6,7,9,11,13 - симфонический оркестр - большой состав;
12 - ансамбль русских народных инструментов (2балалайки, 2 домры, баян)
1,8,10,14 - камерный симфонический оркестр, 11- + ксилофон
5 – хор

Результаты эксперимента показали, что значения τ_e в 74% случаев больше, чем у τ_e, Вч (72 из 92). Параметр τ_e в 24% случаев равен τ_e, Сч, в 6% больше τ_e, Сч, и в 69% меньше τ_e, Сч, и τ_e лишь в 4% случаев равен τ_e, Нч, в 5% больше и 91% меньше τ_e, Нч. Таким образом видно, что значения параметра τ_e (мс) сильно отличаются от его частотно-ограниченных аналогов, в связи с чем вызывает сомнение корректность использования этого параметра.

Более подробный анализ по трем выбранным частотным диапазонам приводит нас к выводу, что общее свойство музыкальных сигналов заключается в их скоротечности в высокочастотном диапазоне. Ясно также, что низкочастотный диапазон отличается большей продолжительностью.

При формировании обрабатываемых wav-файлов использовалась следующая методика. Записи музыкальных коллективов на цифровой носитель производились с помощью микрофонной стереопары (Brul&Kjaer) в относительно заглушенном помещении. Место установки стереопары – 1,5м от края посадки оркестра, ансамбля или хора, высота стойки – два метра. Записывали все произведения целиком, затем сэмплировали временные фрагменты длительностью $20\,сек \quad T \quad 2\,мин$ из разных частей произведений в соответствии со сменой темпа и характера исполнения. Сэмплирование происходило с помощью звуковой станции PRO TOOLS v 6.7 и программы Spectralab v 40.32.08. Два сигнала не складывались в один, поскольку и левый, и правый канал содержал индивидуальную информацию согласно посадке исполнителей.

Выводы. С помощью алгоритмов цифровой обработки акустических сигналов аудиофайлов произведено их распознавание на основе объективного параметра τ_e, характеризующего временную природу случайности процесса возникновения сигнала в музыкальных произведениях. Проанализированы рабочие диапазоны игры различных инструментальных групп по всей частотной шкале и выявлены три основных полосы. Параметр τ_e исследуется как в данных полосах при помощи фильтрации, так и без частотного ограничения. Проведён сравнительный анализ полученных значений. Найдена закономерность, указывающая на скоротечность сигналов в высокочастотном диапазоне.

Библиографический список

1. *Сергиенко А.Б.* Цифровая обработка сигналов / А.Б. Сергиенко. - СПб.: Питер, 2002.
2. *Kuttruff K.H.* Auralization of impulse responses modeled on the basis of Ray-Tracing results // J. Audio Eng. Soc.- 1993. - 41 - P. 876-880.
3. *Pierce A.D.* Acoustics. ASA: New York, 1991.
4. *Блейхут Р.* Быстрые алгоритмы цифровой обработки сигналов / Р. Блейхут. - М.: Мир, 1989.
5. *Солонина А.И.* и др. Основы цифровой обработки сигналов (2-е изд.)./ А.И. Солонина и др. - СПб.: БХВ - Петербург, 2005.
6. *Ando Y.* Architectural Acoustics / Blending sound sources, sound fields, and listeners. Heidelberg: Springer-Verlag, 1998.

Материал поступил в редакцию 30.06.08.

M. A. SUMBATYAN, S. YE. SHEVTSOV

THE ALGORITHM OF DIGITAL PROCESSING FOR ACOUSTIC SIGNALS OF AUDIO FILES AND THEIR RECOGNITION THE BASIS OF SOME OBJECTIVE CRITERIA

There are proposed some algorithms of digital signal processing specific for audio pieces recorded in the form of wav-files. A concept of auto correlation function is introduced, and there are indicated some objective parameters which characterizing the signals of such a type. Some practical aspects of possible aspects of possible classification of various audio pieces are studied.

СУМБАТЯН Межлум Альбертович (р. 1954), заведующий отделом ультразвука НИИМ и ПМ РГУ, профессор кафедры теоретической гидроаэромеханики механико-математического факультета ЮФУ (1999), доктор физико-математических наук (1995). Окончил механико-математический факультет РГУ (1976).
Область научных интересов: акустические методы контроля, исследование фундаментальных вопросов аэродинамики и аэроакустики.
Имеет более 160 опубликованных научных работ.

ШЕВЦОВ Станислав Евгеньевич (р. 1976), главный специалист концертно-театрального центра «Югра-Классик» (г.Ханты-Мансийск). Окончил Екатеринбургскую консерваторию по классу «Звукорежиссура» (1999), а также заочную аспирантуру Санкт-Петербургского университета кино и телевидения по специальности «Акустика» (2003).
Область научных интересов – архитектурная акустика.
Имеет 6 опубликованных научных работ.